

June 1998

## Inside Pandora's Box - The Paper Library: Beyond the Automated Card Catalog

L. Hunter Kevil

University of Missouri, Columbia, [multkevil@showme.missouri.edu](mailto:multkevil@showme.missouri.edu)

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>



Part of the [Library and Information Science Commons](#)

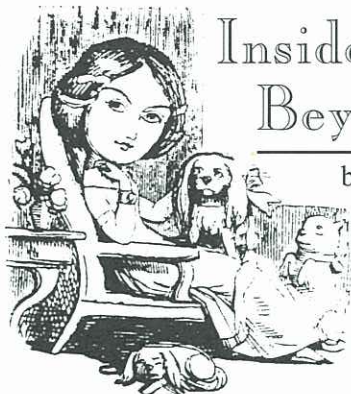
### Recommended Citation

Kevil, L. Hunter (1998) "Inside Pandora's Box - The Paper Library: Beyond the Automated Card Catalog," *Against the Grain*: Vol. 10: Iss. 3, Article 33.

DOI: <https://doi.org/10.7771/2380-176X.2808>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.





# Inside Pandora's Box—The Paper Library: Beyond the Automated Card Catalog

by **L. Hunter Kevil** (Head, Serials Department, Ellis Library, U. of Missouri-Columbia, Columbia, Missouri) <MULKEVIL@showme.missouri.edu>

Column editor: **Jack Montgomery** (Technical Services and Collections Management Librarian, Law Library, University of Missouri-Columbia)  
<MONTGOMERY@law.missouri.edu>

*This essay is a revised version of a concept piece broadcast over the **Cristal-Ed listserv**. (February 1–14th). See <http://www.si.umich.edu/cristaled/discussions.html> for the original and some of the discussion it generated. I would be interested in any of your comments. Thanks. — HK*

I am concerned that without reform and adaptation to contemporary technology our paper libraries as we know them will become increasingly marginalized and eventually little more than book museums. If we are to chart a different course into the future, we must seriously and creatively examine our most deeply held assumptions. I am fully aware that to say the sky may fall down soon is not new news, nor is it widely accepted. The comments below will evoke in many a sense of skepticism I partly share. I offer no silver bullets and no ultimate answers. But since our problems are really social, not technical, even general thoughts such as these are well worth the effort. The dictum that the unexamined life is not worth living applies as much to our professional life as to our personal.

I see four major problems with contemporary library practice. (1) We have a producer-oriented point of view, not a customer- or user-oriented one. (2) The conceptual foundation underpinning our automated systems is basically the same as that developed to handle card catalogs. In other words, libraries have rather uncritically automated their manual procedures and concepts and have failed to take advantage of readily available database-management technology.<sup>1</sup> (3) The system is extraordinarily expensive, largely through complexity and duplication of effort. (4) We have taken insufficient advantage of the opportunities of networked information to pool resources, share data, and create new value for users by working cooperatively.

One of the first questions to be addressed is how libraries can gain a customer focus and change our data structures and our concepts of organization of materials to take full advantage of contemporary database management technology. Without substantial proactive change, a critical mass of users very familiar with accessing and manipulating digital materials will at some point become very intolerant of the anomalies and archaisms of our card-based automated catalogs. Unless libraries transform their local catalogs and assume a leadership role in the emerg-

ing knowledge management systems of their parent organizations, our future may not be pleasant. In short, rather than shoehorn the future to fit past practice and “catalog the Internet,” I suggest libraries should “Internet” their local catalogs to create an open, shared, easy-to-use, cooperatively maintained database management system (DBMS.) This cooperation would enable transferring expenses to those areas offering the greatest return or value to users.

Let me add a bit of perspective. As a search tool, the card catalog is conceptually a set of codified controls over access points and form of entry. The object is to create consistent, usable “headings” at the top of cards so that they can be filed in expected sequences in the author, title, subject, and shelflist catalogs. This inherently manual system relies on staff and users to compare the text strings of the headings with something else, a citation, a book in hand, another card. In time, special concepts and practices incomprehensible to users have arisen, such as uniform entries and titles proper.

This system still governs the operation of our automated catalogs. It appears we cannot get beyond the mental model of catalog cards organized in filing order by their headings. (I too suffer from these mental images—*mea culpa*.) Libraries have used potentially liberating new technology to do little more than what card catalogs could do. Our automated systems are based on a file structure of MARC bibliographic records with indexes of text strings just like headings on cards. The MARC bibliographic format was designed in the 1960's to enable the transmission by magnetic tape of the information on printed catalog cards. MARC authority records perform the same function as their card predecessors.<sup>2</sup> The tools given users have not changed much. The old practices and compromises remain in place, except that now they appear to be obstacles. Our emphasis has always been on treatment of separate bibliographic items (“books”). This means that materials such as serials or conference proceedings, sometimes even books in series or sets, still can be very hard to locate. Our system is one of extraordinary complexity. Over time cataloging rules and local practice have changed. Inconsistency of interpretation, human error, the inability of library staff to perform all the desirable record maintenance, all these problems have created an edifice ready to crumble of its own weight. New search types available in automated systems, such as author-title, keyword, and ISBN, sometimes enable the persistent user to get through the tight control of access points. Yet despite all this ex-

pensive work, our catalogs are not much easier to use than before. Libraries still must maintain large reference staffs to answer largely basic questions. But the persistence required is not worth the bother for many users. It is no wonder that despite some very good efforts with Z39.50, machine searching and processing have not progressed as far as needed.

I want now to look at a possible new way to organize access to our paper resources. These suggestions concern the very simplified logical structure of a possible system independently of any actual implementation. Relational databases have been around more than 25 years. Many of us are familiar with them. This model is interesting for libraries because along with the relational data structure of tables and relationships, we get a built-in query language and something akin to authority control. In this model, data are stored in grid-like two-dimensional-tables. Each table should represent a simple concept or object. The columns in a table represent characteristics (“fields”) and the rows represent field values from actual instances. Each row is identified by a field called the “key.” The key functions as a unique ID similar to an OCLC number. Let us consider an over-simplified model of three tables: authors, titles, and books. The authors’ table could include several fields, including surname, given name, dates, and key. The key field would contain a unique ID assigned by the system. The titles’ table could include fields for title and key and the books’ table fields such as ISBN, publisher catalog number, and key. To establish a relationship between two tables, we enter the key from one into the other. This makes “queries” or searches possible. For example, to find an author’s books, we create a query, which will scan through the books’ table looking for rows identified by the presence of a desired author’s key or unique ID. In this one-to-many relationship, one author can be associated with many books. Many-to-many relationships are possible by creating intermediate tables.<sup>3</sup> Very complex queries are possible and the user can control which fields are returned by the system and the order in which they are displayed. Frequent queries can be saved and easily modified. Each data element is stored only once; unlike with MARC records, there is very minimal or no duplication of data.

These DBMS offer many benefits to libraries. There is no longer a need for unique headings maintained individually by each library, as the system will have been designed to handle multiple identical values such as titles.<sup>4</sup> Global up-

*continued on page 77*



dating is built in; control of headings or entries and authority control in the usual sense are no longer needed. Relational DBMS support verification of data entry, so that an invalid author ID cannot be entered in the books' table, as well as several other kinds of security and integrity and enforcement of "business" rules. Relational integrity, which encompasses data integrity of several types as well as user-defined procedures or business rules, is of crucial importance to libraries. Without relational integrity a shared nationwide system would probably not be possible. In terms of choice of technology, libraries face a three-tiered hierarchy. At present we have an anarchic system in which anyone can edit any field at any time. Mistakes can be created as easily as they are corrected. Most mistakes cannot be detected by machine. Some library systems even permit the editing of indexes. Duplication is built in, with redundant data in our MARC records and duplication of these records in every holding library. The second level is a file-processing system in which integrity and business-rule logic is built into the individual applications comprising the system. For this to work, every single application must be consistent with all the others. Changes in business rules and hardware crashes are problematic for this approach. The third level is represented by the relational model. Data integrity and as many business rules as possible are built into the DBMS itself. Ability to handle rule changes, reliability, performance tuning, recovery from failure are all much en-

hanced.<sup>5</sup> I see no advantage in progressing to the second level rather than simply evolving directly to the third, which is mature and well proven.

Relational databases also excel in supporting queries, particularly when there are simple data types (such as text fields). The important concepts here are relationships by means of unique IDs and a data-centric view rather than a process-centric one. Current library procedures rely on cataloging, the "preprocessing" of data by the **Library of Congress**, bibliographic utilities, and library staff. The approach explored here assumes as a goal a system of well designed tables and relationships, as well as sophisticated tools to handle controlled vocabulary terms such as subjects and tools to handle books whose fulltext has been indexed. This system would then rely on preprocessing of items only to the extent needed to enter "clean" data logically organized. It could exist "on top of" our local MARC databases, with the addition of a few fields in the MARC records and a system of a master look-up table with many mirror sites. Many libraries could share the system. Libraries could move incrementally towards the new system by building extensions to existing authority control structures and moving from a structure of independent local catalogs to a nationally shared network. The general emphasis so far as the user is concerned would be placed on "post-processing," enabling

*continued on page 79*

#### Endnotes

<sup>1</sup> There is an extensive literature describing the problems users experience with library catalogs. Examples include Christine Borgman's articles, "Why are online catalogs hard to use?" and "Why are online catalogs still hard to use?" in *Journal of the American Society for Information Science*, 37:6 (1986) and 47:7 (1996.) This idea has not been generally accepted within the library profession. Perhaps after thousands of searches we have come intuitively to accept the limitations of the typical catalog. Recently a list of humorous questions asked at the reference desk has circulated the Internet. It is interesting to ask why these "howlers" are so funny and by which of our cherished assumptions we consider library users naive for asking. One of these is: "Do you have a list of all the books in the English language?" I find this question both funny and very reasonable. There are in fact many uses for such a list. Why have libraries not yet created such a list or database?

<sup>2</sup> In discussion to my original **Cristal-Ed** contribution, Arlene Taylor pointed out that authority control is evolving toward "access control." In this case information about a person, entity, or topic would be stored in a single place for each library. Since the user would no longer have to match headings, retrievability would be enhanced. What I am proposing is very much along these lines, except that this information would be stored in one place for all libraries, of course with mirror sites and local echoing of pertinent data for performance and other reasons.

<sup>3</sup> A basic explanation of how this works is available in Microsoft's short document, "Designing a database — understanding relational design" (<http://www.microsoft.com/access/productinfo/experttools/urelides/urelides.htm>). A very good six-page introduction to databases for library use is provided by *Biblio Tech Review*, at [http://](http://www.biblio-tech.com/html/databases.html)

[www.biblio-tech.com/html/databases.html](http://www.biblio-tech.com/html/databases.html). The example of the "books" table is of course a gross simplification, shorthand for a complex structure including the concepts of work, instances of a work, containing, contained-in, and related-to. It may well turn out that object database management systems are better suited than relational databases for the relationships that would be required to handle items in series and articles. (Articles are contained in issues, which are contained in one or more volumes, which are contained in a periodical of a certain name and also in a subscription, and so forth.)

<sup>4</sup> For example, consider the journal *Development*. A title search in many OPACs will fail, because the title will be considered to be something like *Development (Cambridge, England)*. Only the persistent user willing to navigate through several title browse screens and make several guesses at the library version of the title will have a chance at success. A machine title search using the Z39.50 protocol would also fail.

<sup>5</sup> A good, brief discussion of these issues can be found in Fabian Pascal, *Understanding relational databases, with examples in SQL-92* (Wiley, 1993).

<sup>6</sup> There should be a strictly enforced rule not to permit enhancing even a single old procedure without also designing a means to evolve to successor procedures in the new system. This is analogous to the dictum never to buy computer equipment that has no upgrade path. For example, the proposed new standard for holdings, Z39.74, recently voted on by NISO members, unhelpfully still recommends a set of user-unfriendly, manually maintained local codes. This is an example to be avoided, as even a few temporary improvements can help defuse momentum for fundamental change.

# puvill

LIBROS S.A.

YOUR BOOKDEALER FOR  
SPANISH AND MEXICAN  
PUBLICATIONS

PROVIDING SERVICES TO  
ALL LIBRARIES

Visit our Web site at:  
<http://www.puvill.com>

#### OUR SERVICES:

- USMARC Bibliographical Information (free of charge)
- Customized Approval Plans
- Periodical Subscriptions
- Serials and Standing Orders
- Firm Orders
- Out of Print Searches
- Rare and Antique Books
- Binding

#### SPAIN

c/ Estany 13, Nave D - 1  
08038 Barcelona - Spain  
Tel. (011 34 3) 298-8960  
Fax (011 34 3) 298-8961  
e-mail: [info@puvill.com](mailto:info@puvill.com)

#### MEXICO

Empresa 109. Col. Mixcoac  
Mexico D F 03910 Mexico  
Tel. (011 52 5) 611-1513  
Fax (011 52 5) 598-4378  
e-mail:  
[74173.1047@compuserve.com](mailto:74173.1047@compuserve.com)

#### USA

264 Derrom Avenue  
Paterson, NJ 07504  
Tel. (973) 279-9054  
Fax (973) 278-1448



## Medic Alert

by Pamela Rose (SUNY at Buffalo)

The new semi-annual *Scientific Review of Alternative Medicine* aims not to "expose" alternative medicine, but rather to present accurate information, according to editor **Wallace Sampson**, Professor of Medicine at **Stanford University**. Proffered as an "antidote" to the **National Institute of**

**Health's Office of Alternative Medicine (OAM)**, the journal aims to make distinctions between "the genuine, the questionable, and the fraudulent." See -- **Holden, Constance**, ed., "New Alternative Medicine Watchdog," *Science*, v.278 (October 24, 1997)

## Mathematical Solution

by Pamela Rose (SUNY at Buffalo)

A new HTML dialect called **Math Markup Language (MathML)** developed by the **World Wide Web Consortium (W3C)** which sets Web standards, will allow mathematical formulas to be written, rather than pasted-in as images, in HTML docu-

ments. **MathML** allows formulas to be understood by screen-reading software, allowing visually impaired people to work with math "in a way that has never been feasible." See -- **Kaiser, Jocelyn**, ed., "Web Math Made Easy," *Science*, v.280 (April 17, 1998), p. 351.

## Finding Frauds

by Pamela Rose (SUNY at Buffalo)

The **National Center for Biotechnology Information (NCBI)** has handed a powerful weapon to fraud police: **PubMed's** push-button "find related articles" feature. Cancer researcher **Marek Wronski** of **Staten Island University Hospital** in New York was able to unearth a trove of thirty

allegedly plagiarized medical papers using NCBI's free, easy-access gate known as **PubMed** to search **Medline**. See -- **Marshall, Eliot**, "Medline Searches Turn Up Cases of Suspected Plagiarism," *Science*, v.279 (January 23, 1998), p. 473-474.

## Inside Pandora's Box

from page 77

the user and in particular user clients and agents to perform virtually unlimited searching.

Let me conclude by listing just a few of the many requirements libraries should demand of any new system, whether brand new or just tweaked. This list is not intended to describe in any detail a possible new system, but to show the variety of considerations a good design will take into account.

- The first and foremost requirement is for simplicity. We librarians are notorious for making our work too complex. The new system should be simpler than the current one. Simplicity usually brings benefits regarding comprehensibility, reasonable cost, a low error rate, and easier training and maintenance. If the new system I am proposing appears to be complex, attribute this to my failure to communicate well, not any inherent difficulty in creating a simpler rather than a more complex system.

- User control of searching — a corollary of the simplicity requirement. Users should be able to search all the data in the catalog, not just the

indexes. This implies that data elements must make sense to the user, that all data should be accurate and "clean," not transformed or altered by libraries. Accurate data requires automatic updates — even if an item in need of updating has already been cataloged "correctly." It also requires that labels, operations, and displays should be readily comprehensible to users. Users for instance should not have to know the distinction between "monographs" and "serials."

- The new system should work in conjunction with the digital library. Users of the system should be able to search for digital materials as easily as paper ones. The design should accommodate progressive steps towards integrating the paper and the digital worlds.

- Libraries should concentrate effort on development of standards for a broad variety of different search tools. Machines as well as humans should be able to use these tools. If these standards extend down to the software component level, they would be very useful in standardizing library software and influencing its development in a positive direction. Lastly, customizable and personal agent-based front-ends and user interfaces are urgently needed.

## Chemistry Online

by Pamela Rose (SUNY at Buffalo)

The **National Institute of Standards and Technology's (NIST's) Online Reference Databases** constitute a rare source of free information on chemicals. NIST experts scrupulously assess,

reference, and update the data, which includes everything from melting points to half-lives. See -- **Kaiser, Jocelyn**, ed., "Hot Picks," *Science*, v.280 (April 17, 1998), p. 351.

## TOC News

by Pamela Rose (SUNY at Buffalo)

The **American Chemical Society (ACS)** will unveil free access to TOC for all twenty-six of its journals back to January 1996, at:

[pubs.acs.org](http://pubs.acs.org). See -- **Kaiser, Jocelyn**, ed., "Straight Dope on Chemicals," *Science*, v.280 (April 17, 1998), p.351.

## Gatekeeping

by Sandra Beehler (Old Dominion University)

When a hacker named **Eugene Kashpureff** intercepted new domain name registrants on their way to the root server, he brought to attention the controversy surrounding control of the Internet's domain name system (DNS). Most agree that controlling the DNS is tantamount to controlling the Internet itself. The contract held by the current official keeper of domain names, **Network Services, Inc.**, is running out in September, and the fight is on to control this lucrative system. A variety of so-

lutions have been proposed, including privatization, globalization, and open competition. This article outlines the proposed solutions while it tells the tale of Kashpureff's "stunt." See — "Whose Internet is it, anyway?" *Wired*, v.6 (April 1998), p. 172.



- The new system should be designed to support local collection analysis, acquisitions, serials, interlibrary lending, and document delivery. It should at least support prepublication title and publisher's catalog number or other unique ID similar to the **Digital Object Identifier**. It should recognize subscriptions and the issues they cover as well as licenses for digital materials.

- The new system should increase library productivity and generate savings from old activities that could be used to pay for new ones. One way to do this is to standardize the objects libraries work on so that human editing and interpreting of local text fields is minimized or eliminated. Right now every library reviews the same **Library of Congress** cataloging copy for misspellings and other errors. If this duplication of effort were eliminated, library staff would be freed to do other, more important work, such as maintaining the new cooperative system, analyzing whatever is analyzable, extending item control to the article level, or digitizing their unique and important items.

- The new system should be designed from the ground up with full support for sharing and

*continued on page 80*



# Group Therapy — Prepaying Rush Orders

Column Editor, **Rosann Bazirjian** (Florida State University)

Hey y'all out there! Do you have any gripes? Come to your therapist! Try <rbazirji@mailers.fsu.edu> or FAX 904-644-5170.

**GRIPE:**  
(Submitted by **Robert Schlabbach**, Library Business Manager, The College of New Jersey)

We recently ordered from **Amazon.com** for the first time. We needed a rush reserve that was OSI at the publisher and the requester told us it was available from Amazon.com. Sure enough, their Website indicated that it was available. They had one copy in stock, but our library doesn't have a credit card and Amazon.com wouldn't hold the copy for us. We did a rush prepay, but it will still take three weeks for our check to arrive and clear. Wouldn't it be nice if library vendors included inventory on their online databases, i.e., **BNAm's NTO** and **YBP's GOBI**? Sure, they're reluctant, but it would help immensely in locating copies of OP and OSI titles. I think for now we'll probably establish a deposit account with Amazon.com.

**RESPONSE:**  
(Submitted by **Bob Nardini**, Regional Vice Pres., Collection Management and Development Group, Yankee Book Peddler)

Sometimes when a vendor steps into the batter's box, as it were, to answer a question, a librarian will serve up a homerun ball that you can knock into the seats, and then trot around the bases looking like Mickey Mantle.

But not every inning. Sometimes what you get is a fastball you can barely see. The best you can do is to foul it off. This question is a hard fastball, one no vendor is going to hit for a homerun—not, at least, any vendor oriented toward academic libraries, where our customers call on us to have the capability to supply a very wide range of titles — many of them esoteric,

difficult to locate, and bound to be needed by only a few libraries — and to fill orders accurately and in a customized manner according to individual library specifications almost limitless in their variety.

That's a very different mandate than the one guiding vendors oriented toward public libraries and retail stores where the ability to supply a more predictable range of titles very quickly — from local inventory — wins and holds customers.

Academic vendors may in fact carry substantial inventories, but it would be extremely expensive to compete directly with one another on the basis of what was on the shelves at a given moment.

Not only would the cost of added inventory have quite a price tag — one that not all vendors could afford — but development costs for effective real-time inventory windows would themselves be substantial, and would mean that development or expenditures in other areas would have to wait.

We agree that it would be nice to include inventory on our databases. But we're not certain, knowing that costs would be high, that this should be a priority. Not, especially, since it is already possible for a librarian to call or email any academic vendor about a book that might be needed — in just a bit more time than it would take to search a database.

Not a homerun, but this is spring training, after all.

**RESPONSE:**  
(Submitted by **Matt Nauman**, Marketing Manager, Blackwell North America)

When Rosann asked me to respond to the suggestion that library vendors include inventory on their online databases, I thought perhaps it was time to go on spring break. Then for some

reason valor overcame discretion. Academic booksellers see this question with some regularity and it deserves a response.

The simple answer here is still the best. While some scholarly carry substantial inventories, they are quick turn-around inventories. New titles and firm orders are based on existing customer needs. They arrive, are invoiced, and sent out to customers as quickly as possible. These are the bulk of the books in any warehouse and it would be difficult to list them as inventory on hand.

The books that are held in inventory over a longer period of time frequently come from two main sources: returns and actual approval stock buying. Like books for approvals and firm orders, returns are generally turned around as quickly as possible. These books are used to fill new orders or returned to publishers. Such also has been the case with stock quantities that buyers bring in to augment approval buying. The quantities are used to fill orders based on new title announcement slips or firm orders. The quantities are usually modest and are not meant to be held for a long period of time. I have wondered about the possibility of designating such titles in our database as "stock quantity available."

What's in inventory if one's inventory is fast-turning is a tough question. On the other hand, booksellers understand how useful this information can be. I hope that at some point we will be able to address it. Newer technology may enable vendors to solve the problem in the near future. For now, several library booksellers offer inventory check services through customer service departments. If you phone, fax, or email, vendors will provide inventory status reports. 🐾

## Inside Pandora's Box from page 79

cooperation. For guaranteed fast response time libraries could locally maintain full or partial data for their own holdings and rely on network searches for outside resources. This implies a national or worldwide search by ID and extends the notion of user searching far beyond the local catalog. This has many important benefits. For example, this would for the first time enable the user to determine positively that a given library does not hold a given book or serial (or article.) These IDs would help prevent the creation of duplicate records, something current systems

cannot do well. With a shared national system it would be possible to enable searching for many items that are not present in current MARC records or library indexes, such as periodical articles and chapters and contributions in books.

•• The new system should be designed so that it can be built automatically from existing MARC records with a minimum of manual work. Deciding on the structure of the tables of the new system and the mapping of MARC fields will be interesting and challenging work. Once this has been accomplished, Internet spiders could read through LC and local catalogs, populating the tables with data. Duplicate resolution would form an interesting challenge, but need only be

done once and should be shared. Each cooperating library would link its call numbers and all other unique local information. These tables of local data would be maintained by libraries in a distributed arrangement. The processing or "cataloging" of new paper works could be shared with publishers and book distributors. Libraries could continue to use their MARC catalogs as long as they wished. It may be possible to offer users the old or the new system with their choice of user interface. The example of Internet search engines shows that a new national catalog could be created. If libraries do not do this, I fear a competitor, such as a company in the business of outsourcing libraries, will. 🐾